DECEMBER 2008 | ISSUE 1

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

# eProtein

## *Letter from the Director*

Dear Colleagues,

Since the Clinical Proteomic Technologies for Cancer (CPTC) initiative was launched just over two years ago, tremendous progress has been made towards reducing the layers of systems variability that plague candidate protein biomarker discovery. This is not surprising given that we have an amazing scientific network—the best minds in proteomics—representing nearly 50 federal, academic, and private sector organizations who are deeply committed to open and collaborative science for the sake of the entire cancer community. A great deal of work still remains before us, but we are confident that the investment that the National Cancer Institute (NCI) is making in proteomics today will pave the road for clinical translation tomorrow.

CPTC is pleased to launch this online quarterly newsletter, *eProtein*, as a way for the proteomics community to keep abreast of the many exciting achievements being made within the CPTC programs as well as by individual colleagues all across the nation. We are committed to ensuring the success of this important initiative because together we are building the foundation for clinical cancer proteomics. ■

## *The 2nd Annual CPTC Meeting, Cambridge, Mass.*

CPTC held its second annual meeting in Cambridge, Mass. on October 28–29, 2008, bringing together more than 200 participants representing the full gamut of scientific fields that contribute to the initiative's mission to review the technological progress made over the previous year.

Giving a sense of the links between CPTC and other technology focused initiatives supported by NCI, the

*continued on page 2*

## *New Antibody Portal Puts Well-Characterized Antibodies at Researchers' Fingertips*

A key challenge for proteomic researchers is seeking out and acquiring high-quality, well-characterized monoclonal antibodies. While numerous commercial reagent suppliers make antibodies available for research, their antibodies tend to be expensive and may or may not be extensively characterized. Thus, researchers can at times be left guessing whether an antibody

*continued on page 3*

**A Clinical Proteomic Technologies for Cancer initiative publication that builds connections throughout the proteomics community**

## In This Issue

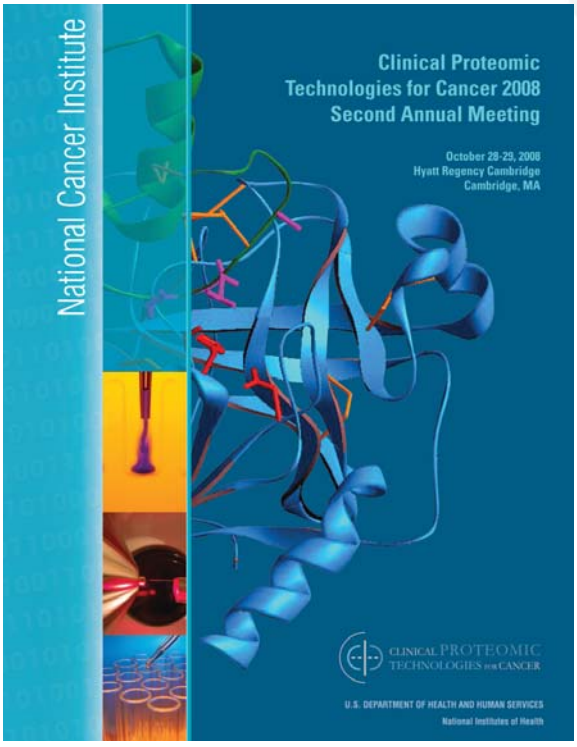
# *The 2nd Annual CPTC Meeting, Cambridge, Mass.*

*(continued from cover)*

first day of the meeting was held jointly with members of NCI's Innovative Molecular Analysis Technologies (IMAT) program.[1] Several talks featured technologies and techniques developed by IMAT-supported investigators that have subsequently been applied to projects supported by CPTC, highlighting the importance of integrated technology development in cancer proteomics research in particular and in cancer research in general. The meeting also included talks and posters featuring research conducted through CPTC's three components: the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program, Advanced Proteomic Platforms and Computational Sciences, and the Proteomic Reagents and Resources Core *(see New Antibody Portal Puts Well-Characterized Antibodies at Researchers' Fingertips, cover and page 3)*. Both days featured keynote addresses by researchers speaking on their experiences in integrated research. David Altshuler, M.D., Ph.D., a founding member of the Eli M. and Edythe L. Broad Institute of MIT and Harvard and Director of the Institute's Program in Medical and Population Genetics, spoke of the lessons learned from conducting large-scale genomics research and how those lessons could apply to large-scale proteomics. In particular, he noted ways of avoiding pitfalls in validating variations, such as early development of robust, comprehensive, and scalable tools for determining systematic associations—an issue that, he noted, CPTC is well on its way to addressing by focusing on technology development up front. Altshuler concluded his remarks by reminding attendees of the importance of grounding new discoveries in human biology before jumping to conclusions about their importance.

The keynote address on the second day, given by Vamsi Mootha, M.D., of the Broad Institute and Massachusetts General Hospital, focused on integrative genomic, proteomic, and metabolomic research on mitochondrial diseases. The mitochondrial proteome has not yet been fully defined, but it may contain between 1,200 and 1,500 proteins, only 13 of which have been associated with genes found in the mitochondrial genome; the rest are encoded by nuclear genes. Mootha's talk outlined his work to develop a mitochondrial protein catalog, called MitoCarta, which currently contains 1,098 mitochondrial proteins. With this information in hand, he has started probing the ancestry of numerous mitochondrial proteins and applying that knowledge clinically to explore rare familial diseases caused by breakdowns in mitochondrial respiration oxidative phosphorylation.

In his closing remarks, CPTC Director Henry Rodriguez, Ph.D., M.B.A., noted that the initiative had produced some very good outputs since its launch two years ago. Rodriguez also mentioned that while there had been a learning curve associated with the initiative, they had shown that team-based science could be very successful, and that the steps that had been undertaken thus far had laid the groundwork for CPTC's future success.




> *"CPTC is definitely on track to meeting the goal of optimizing cancer proteomics research."*
>
> *– Moyez Dharsee*
> *Director, Informatics*
> *Ontario Cancer Biomarker Network*

*The third CPTC Annual Meeting will be held on October 5-7, 2009, in Bethesda, Md. Information on the meeting will be posted at http://proteomics.cancer.gov as it becomes available.* ■

[1] To learn more about IMAT, visit the program Web site at *http://imat.cancer.gov*.

# *New Antibody Portal Puts Well-Characterized Antibodies at Researchers' Fingertips*

*(continued from cover)*

appropriate for their experimental platform is available for their studies.

At the CPTC Annual Meeting in October 2008, the initiative's Proteomic Reagents and Resources Core announced the launch of the Reagents Data Portal (http://cpti.abcc.ncifcrf.gov/), a Web-based service open to the scientific community that helps scientists search for and access antibodies from a collection of highly characterized mouse monoclonal antibodies for cancer-associated proteins. According to Tara Hiltke, Ph.D., a Program Manager at CPTC, "Users can perform a keyword or alphabetical search to look up a protein, see the antibodies available for it, see the characterization information for those antibodies, choose the one that best fits their needs, and seamlessly order it from the repository at the University of Iowa's Developmental Studies Hybridoma Bank [DSHB]."

Antibodies in the collection are being generated against 1,261 tumor-associated proteins listed by Anderson and Polanski in 2006 as part of a collaboration between CPTC and several laboratories and companies.[1] Argonne National Laboratory (ANL) clones and expresses each of the proteins, which are then provided to private sector partners contracted for antibody generation through requests for proposals. Each contractor receives 40 proteins from ANL and generates 10 monoclonal IgG antibodies for each, ultimately submitting three of the 10 for in-depth characterization at four collaborating centers: NCI-Frederick, NCI's Center for Cancer Research Tissue Array Research Program (Bethesda, Md.), the Harvard Institute of Proteomics (Cambridge, Mass.), and the Human Protein Atlas at KTH Royal Institute of Technology (Stockholm, Sweden).

> *"We're trying to make as much information available to the community as possible on each antibody so that members can reproduce, trust, and use these antibodies."*
>
> *– Tara Hiltke, Ph.D.*
> *CPTC Program Manager*

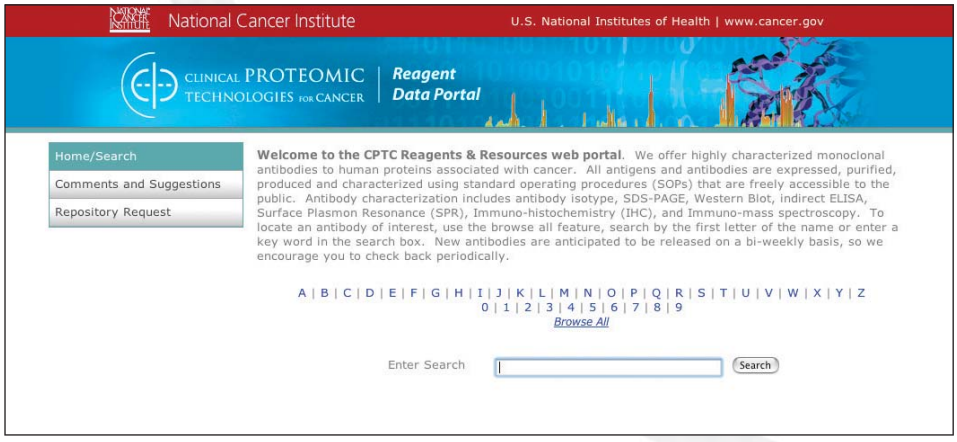
Each antibody is analyzed using:

- ELISA
- Immunohistochemistry
- Immuno-mass spectrometry
- Nucleic acid programmable protein arrays
- SDS-PAGE
- Surface plasmon resonance
- Tissue arrays
- Western blot

All of the antibodies and their associated hybridomas are deposited in the DSHB; once characterization is complete, they are made available to the public at a significantly discounted price.

A key advantage of the Portal is the depth of data available. All of the associated characterization information is available for each antibody, as is detailed information on each protein (i.e., sequence, molecular weight, isoelectric point, alternate names, accession numbers, DNA source, and expression system). In addition, detailed standard operating procedures are posted for protein and antibody generation and characterization analyses. "We're trying to make as much information available to the community as possible on each antibody so that members can reproduce, trust, and use these antibodies, and so that they can serve a wide variety of applications," said Hiltke. ■

[1] Polanski M and Anderson NL. A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights*. 2006;2:1-48.



# *Facilitating Data Sharing and Release in Proteomics*

*[Ed. The following is a condensed version of an editorial written for the Journal of Proteome Research.[1]]*

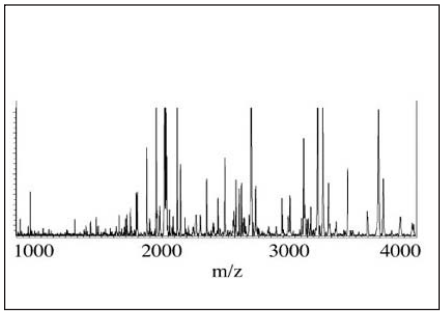
Data sharing is standard practice among members of the genomics community, based on principles developed at a 1996 gathering in Bermuda and ultimately endorsed by all major parties in the Human Genome Project.[2] The widespread sharing of prepublication sequence data greatly accelerated the pace of genomic discovery. However, similar policies do not exist for proteomics research, a state of affairs currently seen as a significant obstacle to progress in the field.

> *Their task: to begin defining policies and practices that would govern and facilitate the release of proteomic data into the public domain along the Bermuda model.*

In August 2008, the NCI sponsored a summit in Amsterdam for members of the international proteomics community, including representatives from funding agencies, journals, and academic research centers. Their task: to begin defining policies and practices that would govern and facilitate the release of proteomic data into the public domain along the Bermuda model. The summit's participants addressed the following questions:

## When should data be released?

Participants agreed that the timing of data release should be governed by the type of project. Investigators taking part in community resource projects should be required to release data once they are produced. Investigators working on individual projects, on the other hand, should release data upon publication in a peer-reviewed journal.




## What types of data should be released, and what kinds of metrics should be used to define data quality?

Participants agreed that high-quality, well-annotated raw data (for mass spectrometry and protein/affinity array data) would be the most reliable interchange format for data repositories. Metadata, information on data quality, and identification quality control will all be critical as well. Accessing these data would require development of the proper infrastructure (i.e., community supported standardized formats, controlled vocabularies and ontologies, minimal reporting requirements, and publicly available online repositories). Central repositories should develop their own thresholds for data quality metrics, in a coordinated manner with users and one another, to ensure interoperability.

To fuel progress in proteomics research, data sharing cannot be voluntary; rather, it is up to scientists, journals, and funding agencies to take the necessary steps to ensure that all parties adhere to the standards for data release, ideally within a framework of tripartite responsibility akin to that created for genomics research.[3] Central repositories, for their part, should clearly define minimum submission requirements, encourage rich annotation, and develop seamless submission procedures.

A white paper based on the discussions of the summit is forthcoming. ■

[1] Rodriguez H. International summit on proteomics data release and sharing policy. *J Proteome Res.* 2008 Oct 7. [Epub ahead of print]

[2] Policies on Release of Human Genomic Sequence Data. US Department of Energy, Human Genome Project. *www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml.* Accessed 22 October, 2008.

[3] Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. The Wellcome Trust. *http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf.* January 2003.

# *SBIR Program Helps Integrate Technology Development Efforts by CPTC and Small Businesses*

Ready access to high-quality, standardized reagents is of great importance if the proteomics community is to catalyze biomarker discovery for reducing the burden of cancer. Numerous small businesses design and develop proteomic technologies for the accurate and powerful measurement of proteins and other biomolecules related to disease. Without well-characterized reagents, however, it is impossible to translate such platforms into products and services that could be used effectively by the cancer community.

One of CPTC's three component programs, the Proteomic Reagents and Resources Core, is tasked with providing the cancer community with the tools necessary to overcome technological and methodological barriers to developing and providing such reagents. To maximize the Core's capabilities and impact, CPTC is partnering with the biotechnology industry via the NCI's Small Business Innovation Research (SBIR) Program, a contract mechanism that supports early-stage research and development by small businesses. Through the SBIR program, CPTC aims to integrate its efforts with those of the biotechnology industry by encouraging and enabling companies developing proteomic technologies and platforms to adopt standardized, well-characterized reagents—including high-quality proteins and validated capture reagents (e.g., antibodies)—in the commercialization of new tools and kits for the cancer community.

CPTC has already awarded contracts based on SBIR requests for proposals released in 2007 and 2008 on such topics as "Development of Clinical Automated Multiplex Affinity Capture Technology for Detecting Low Abundance Cancer-related Proteins/Peptides" and "Advances in Protein Expression of Post-Translationally Modified Cancer Related Proteins." For instance, SBIR-awardee Rules-Based Medicine Inc. is customizing CPTC-developed reagents for a quantitative, automated, Luminex-based 50-plexed immunoassay for the rapid detection of low abundance cancer-related proteins.

For fiscal year 2009, CPTC sought proposals on "Novel Antibody Epitope Mapping Technologies," "Development of Novel Protein Expression Technologies for Glycosylated Cancer Related Proteins," and "Peptide Aptamers: New Tools to Capture and Study Protein Interactions in Lieu of Immunological Reagents." Contract awards for these topics are anticipated to be announced in the summer of 2009. ■

*Through the SBIR program, CPTC aims to integrate its efforts with those of the biotechnology industry.*

## *2007 SBIR Contract Recipients*

**Development of Clinical Automated Multiplex Affinity Capture Technology for Detecting Low Abundance Cancer-related Proteins/Peptides**

| | |
|---|---|
| Meso Scale Diagnostics | Automated Multi-Array Platform for Cancer Biomarkers |
| Sequenom, Inc. | Sensitive Protein Detection Combining Mass Spectrometry |
| Quadraspec, Inc. | Highest Sensitivity Cancer Marker Array on Quadraspec's Bio-CD Platform |
| Rules-Based Medicine, Inc. | Automated Multiplexed Immunoassays for Rapid Quantification of Low Abundance Cancer-Related Proteins |

**Development of Alternative Affinity Capture Reagents for Cancer Proteomics Research**

| | |
|---|---|
| Allele Biotechnology & Pharmaceuticals | Yeast Single Chain Antibodies as Capture Reagents |
| Accacia International, Inc. | High-Throughput of Aptamers against Cancer Biomarkers |

## *2008 SBIR Contract Recipients*

**Advances in Protein Expression of Post-Translationally Modified Cancer Related Proteins**

| | |
|---|---|
| Rana Biosciences, Inc. | A Cell-Free System for High Yield Phosphoprotein Synthesis |

**Development of Clinical Quantitative Multiplex High-Throughput Mass Spectrometric Immunoassay for Detecting Low Abundance Cancer Related Proteins/Peptides in Bodily Fluids**

| | |
|---|---|
| Intrinsic Bioprobes, Inc. | Multiplex Mass Spectrometric Immunoassays |
| Predictive Physiology and Medicine, Inc. | Immunoaffinity Capture Coupled with Ion Mobility |

# *Researcher Spotlight: David Tabb, Ph.D.*

## *Making Sense of the Complexity with User-Friendly Tools*

***Starting in the next issue, the Researcher Spotlight section will highlight investigators from the CPTAC network as well as the Advanced Proteomic Platforms and Computational Sciences program.***

David Tabb, Ph.D.

A bioinformatics researcher at Vanderbilt University, David Tabb, Ph.D., views his field from an integrated point of view, based on an underlying philosophy that software tools should be designed both for high performance and for comprehensibility by end-users.

Though he began his career by focusing his efforts on characterizing peptide fragmentation, Tabb now devotes his work to improving peptide identification through both database search and sequence tagging. He aims to integrate the two approaches to maximize the biological information produced from proteomics experiments.

Tabb initially joined CPTC as part of its Advanced Proteomic Platforms and Computational Sciences program. "CPTC was looking for algorithms to make sense of proteomic data, the area I find most interesting." He has since chaired the CPTAC Bioinformatics Working Group. "Funding from CPTC enables my team to develop research tools, while the Vanderbilt CPTAC funding enables us to apply and refine them for others," he says.

At all times, Tabb tries to keep research utility at the forefront of his design and development processes. "Too often, bioinformaticists view data only as data," he notes, "without looking at the biological picture or at the real need of laboratory biologists to understand how an algorithm produces a result." To that end, Tabb and his team work closely with the Vanderbilt CPTAC team, taking laboratory researchers' needs and perspectives into account. "My goal is to validate the tools my team and I are developing across multiple platforms and laboratories, so they can have the broadest biological utility."

For instance, because CPTAC's inter-laboratory experiments produce huge data sets that can be analyzed in multiple ways, the Tabb laboratory has emphasized algorithms that scale well for complex experiments instead of one small experiment at a time. Also, they try to address questions of usability. "Most tools are designed with a command line interface," Tabb says, "which few biologists are comfortable with." Thus, Tabb's group is working to add more user-friendly graphical interfaces to their algorithms. He is also building in connections to other software tools. The Tabb group is broadening the utility of IDPicker—their program for assembling protein sequences from raw peptide database identifications—by adding connections to spectrum review and sequence coverage tools and by incorporating the ability to accept identifications from any database search tool that can output pepXML-formatted data.

With the CPTAC Bioinformatics Working Group, Tabb is also working on methods to make CPTAC data available through the cancer Biomedical Informatics Grid (caBIG ), an NCI initiative intended to develop a nationwide information technology infrastructure for cancer research. "CPTAC program data will be made available to the scientific community via caBIG®," Tabb explains. "This presents its own challenges, as there are few tools or standards available in caBIG® for proteomics. We are working hard, though, to ensure that raw data will be available to anyone who wants them." ■

> *"My goal is to validate the tools my team and I are developing across multiple platforms and laboratories, so they can have the broadest biological utility."*
>
> *– David Tabb, Ph.D.*
> *Assistant Professor of Biomedical Informatics*
> *Vanderbilt University School of Medicine*

# *Researcher Spotlight: Xiaolian Gao, Ph.D.*

## *Visions of High-Throughput Clinical Proteomics*

Xiaolian Gao, Ph.D.

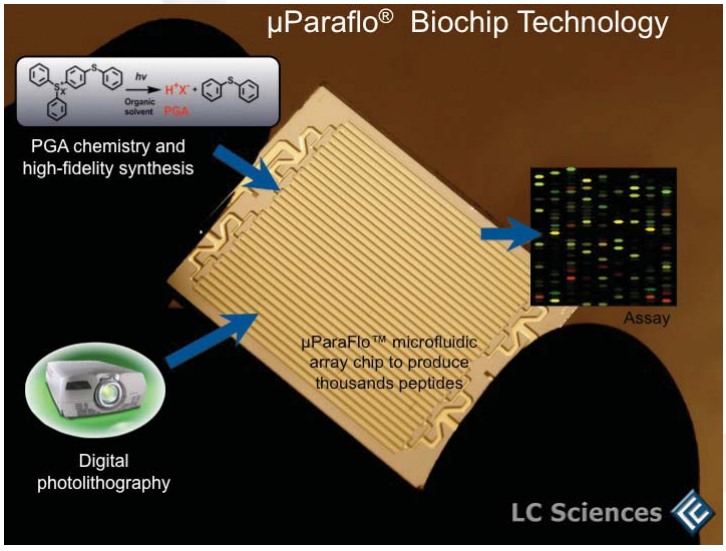
Small, inexpensive, fast, and able to assess the activity, structure, and interactions of thousands of genes at a time, DNA microarrays are a mainstay of genomics research. Protein microarrays have the potential to bring the same high-throughput analytical power to proteomics. "The microarray format is very familiar in the genomics world," says the University of Houston's Xiaolian Gao, Ph.D., "but it is still new in proteomics. However, the community is quickly recognizing its value as a means of running high-throughput experiments in ways that keep false positive rates low." Gao is contributing to the format's value by bringing array production and analysis together in the form of a microfabricated, high-density, addressable microarray platform being developed through CPTC's Advanced Proteomic Platforms and Computational Sciences program.

Gao's microarray technology is based on *in situ* peptide synthesis using a combination of microfluidics and methods akin to the photolithography techniques used to construct computer microchips. With these techniques, she can reproducibly achieve far higher peptide densities than most chips—3,000 to 4,000 peptides per 1-cm$^2$ array, as opposed to a few hundred peptides—using only microliter amounts of reagents per array.

The chips, which can be analyzed using standard DNA microarray fluorescence scanners, are also addressable. "The solution flow through our microfluidic chip is very much like chromatography," she says. "By controlling flow rate and temperature, we can run parallel synthesis and assays on a single chip. We can place individual peptide sequences in replicates and control and reference peptides at regular locations to minimize false positive measurements and increase stringency." Because of its flexibility, this array technology could bring new resolution to such experimental techniques as epitope mapping. "Generally, if you have a new antibody and want to find its binding site, you design and synthesize an epitope chip with 8- to 10-mer peptides derived from the antibody's target protein, add the antibody, and look for where it binds," said Gao. "With addressable *in situ* synthesis, we can both identify an antibody's core epitope sequence at a single amino acid resolution and quantify its binding constant. One microarray chip can contain as many titrations as 41 96-microwell plates, allowing us to compare the specific binding of several antibodies simultaneously."

Gao is using the chips to study cancer-related signal transduction cascades by probing the interactions of phosphoprotein-binding proteins. "We can synthesize chips with peptide substrates that only recognize particular phosphoprotein-binding domains and develop signaling protein expression profiles for biospecimens such as serum or lysed cells. [The University of Minnesota's] Tongbin Li, Ph.D., and I are developing a database of these interactions that we hope will serve as the basis for peptide chip applications and computational systems biology efforts."

Gao envisions translating her microarray platform into an inexpensive, high-throughput system for clinical proteomics. "Ideally, a clinical laboratory would have a microfluidic chip analysis station, with which they could routinely take almost any patient specimen—blood, plasma, serum, digested tissue, maybe even urine—and look for diagnostic biomarkers or protein profiles."■



CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

## Upcoming Events

**February 22-25, 2009**
US HUPO
5th Annual Conference
San Diego, CA

**April 18-22, 2009**
American Association of Cancer Research (AACR)
100th Annual Meeting
Denver, CO

**October 5-7, 2009**
Clinical Proteomic Technologies for Cancer
3rd Annual Meeting
Hyatt Regency Bethesda
Bethesda, MD

**For a full list of upcoming events, visit http://proteomics.cancer.gov/mediacenter/events.**

## In the Next Issue of *eProtein*:

**A Success Story:** The Collaboration between CPTC and NCI's Center to Reduce Health Disparities for the Development of Training Opportunities in Emerging Technologies – Clinical Proteomics

**Community Outreach:** How Engagement with Patient Advocates Enriches the CPTC Program

**The NCI Clinical Proteomic Technologies for Cancer initiative seeks to foster the building of an integrated foundation of proteomic technologies, data, reagents and reference materials, and analysis systems to systematically advance the application of protein science to accelerate discovery and clinical research in cancer.**

## Reagents Data Portal

*Newly Released Antigens and Antibodies*

| Antigen | Antibody |
|---|---|
| Interleukin 18 | CPTC-IL18-2 |
| Metastasin 100 calcium-binding protein A4 (Calvasculin) | CPTC-S100A4-1<br>CPTC-S100A4-2<br>CPTC-S100A4-3 |
| Nucleoside Diphosphate Kinase B | CPTC-NME2-1<br>CPTC-NME2-2<br>CPTC-NME2-3 |
| Ras-related C3 botulinum toxin substrate 1 (rho family small GTP binding protein Rac1) | CPTC-RAC1-1<br>CPTC-RAC1-2 |

## Contact Information

For more information about the CTPC, please visit *http://proteomics.cancer.gov*, or contact us at:

National Cancer Institute
Office of Technology & Industrial Relations
ATTN: Clinical Proteomic Technologies for Cancer
31 Center Drive, MSC 2580
Bethesda, MD 20892-2580
Email: cancer.proteomics@mail.nih.gov